

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Potential Applications and Challenges of Metagenomics in Human Viral Infections

---

Prudhvi Lal Bhukya and Renuka Nawadkar

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75023>

---

## Abstract

Complex association of human host and pathogenic viruses makes a necessity to understand the overall host and virus interaction network. Identification of virus population and its systematic classification will help in understanding the viral association with the disease outcome. Metagenomics is a recently developing approach for the detection of pathogens in the samples with precise interpretation in a short period of time. Metagenomic approaches have been employed for studying the predominance or spread of the virus within a particular locality and nature of virus during infection. Metagenomics is basically a collective approach of lab-based techniques and in-silico methods for identification of pathogenic viruses without culturing them in specific aseptic conditions. Lack of unique conserved genes in viruses has made metagenomics study difficult in this juncture. Other challenges in the field of metagenomics are like cellular DNA contamination, free environmental DNA contamination and continuous evolution of viruses. Recent studies have shed light on the advancement of this field in virus identification and characterization however still needs further investigations to overcome the challenges. Current chapter focuses on the application and challenges faced in metagenomic analysis of human viral infections.

**Keywords:** metagenomics, viral metagenomics, gastrointestinal infections, applications of metagenomics

---

## 1. Viruses

In Latin, the term virus means toxin, virus are obligate intracellular parasites with RNA or DNA as a genetic material. They vary in size from ~20 nm to ~1.5  $\mu$ m and simple machinery. Viruses can't be able to replicate themselves as they are intracellular parasites and require susceptible host

---

for their propagation. Extracellular viral particles are noninfectious in nature. They can infect a wide range of hosts including plants, bacteria, fungi, algae, protozoa, vertebrate or non-vertebrate animals. In nature, around  $1 \times 10^{31}$  number of different viruses are present. The number itself suggests the diversity of viruses in nature. They play a very important role such as an increase in diversity via horizontal gene transfer in hosts, and nutrient recycling [1]. Report from Hooda et al. showed the abundance of viruses in nature is around 1000 times more than observed via cell culture dependent technique [1, 2]. This suggests the large pool of viruses is still unknown, only around 219 pathogenic viruses have been yet identified [2, 3].

## 2. Role in pathogenesis

**Human viruses:** More than 200 viruses are known to infect humans and number is increasing with time, but the diversity of viruses suggests a huge number of viruses still unknown. In humans, yellow fever virus was the first pathogenic virus discovered in 1901. 1900 was the era of human virus discovery and most of the common pathogenic viruses studied during this time. In current scenario, two out of three infection causing organisms are viruses [4] and known to cause a variety of disease ranging from normal acute infections such as common cold, flu, and gastroenteritis to deadly diseases such as Hantavirus pulmonary syndrome (Hantavirus), AIDS (HIV) ebolavirus disease (ebolavirus). Recent outbreaks of viruses show the emergence of previously known viruses with modified virulence properties.

## 3. Human gut and viral infection

For decades human gut-associated pathogenic viruses are known for many gastrointestinal diseases as gastroenteritis. Following are the main group of viruses has been identified. Rotavirus, adenovirus (serotype 40 and 41), astrovirus, calicivirus, norovirus, torovirus, herpesviruses, coxsackieviruses, human papillomaviruses [3], Norwalk-like viruses, coronaviruses, picornaviruses, Sapporo-like viruses [4, 5]. They infect epithelial cell linings, mucosal linings of the stomach and small intestine, a specific portion of epithelium in the intestine. Depending upon the infection type observed, different samples are used for detection of the infectious agent. In general, feces sample used for general microbiological examination during gut-associated infection [6, 7]. Apart from feces, gastric biopsy, gastric juice, saliva [8, 9] duodenal fluid, cotton swabs [5] are collected. These samples are very essential for diagnosis as they directly contain the pathogen.

## 4. Methods for diagnosis and virus identification

### 4.1. Traditional methods

Since viruses are extracellular inert particles they need to be propagated into on susceptible host or host cells for their growth. Initially, viruses were cultured in vitro with the help

of embryonated eggs or laboratory animals. Discovery of tissue culture technique in the 1900s provides an indispensable tool for in vitro virus culture. Tissue culture technique has been then recognized as a “gold standards” for virus discovery. Major advantages of using tissue culture technique for virus identification are an amplification of viruses, characterization of the virus, functional studies, drug targeting, and genome extraction. Due to authentic results and sensitivity of the technique, tissue culture-based techniques are still in use for virus discovery, as well as immune responses study, altered gene expression and characterization of viruses. Successful use of tissue culture technique in virus identification depends on crucial steps involved such as collection of a sample from high titer area of the body, immediate transport of sample, sample processing and selection of appropriate cell line [10]. The major defects of traditional method for virus identification are difficulties in identification of susceptible cell line, time-consuming and laborious in nature [10]. Moreover, culture-based virus identification is further succeeded with the evolution of new scientific techniques and modification in existing techniques. Shell vials with centrifugation, PRE-CFE stain technique, immune-based techniques, e.g., ELISA, agglutination, precipitation, flocculation, microscopy-based techniques, reduced the time of virus identification but compromising sensitivity.

## 4.2. Molecular methods

Gradually field of virology shifted their particles toward molecular biology methods. Together, traditional culture-based methods and molecular biology techniques are used hand in hand for studying virus associated samples [11]. Broadly molecular biology methods are of two types: sequence dependent and sequence independent. Both the methods have proven its usefulness; many viruses have been identified using these techniques.

1. **Sequence-dependent method:** These techniques are most sensitive molecular biology techniques; it can amplify selective DNA from mixed samples [12]. Since the time of discovery of PCR, it has opened the door for many other variations of PCR for multiple gene modulations. The basic backbone of molecular biology PCR is, it has been used in several approaches such as for sequencing of known viruses depending on similarly in sequence in DNA or consensus sequence of previously known viruses, RFLP and diagnostic purposes [13–16]. Another technique, microarray introduced in 1995, it is used mainly for gene expression studies, used in gene profiling, usually in infected samples. Two methods have been used for discovery of new viruses, taxa, gammaretroviruses and xenotropic murine leukemia virus, SARS-CoV are few best examples [17]. The subsequent studies were unable to reproduce the earlier results [6, 7].
2. **Sequence-independent method:** This approach is independent of prior knowledge of virus genome sequence. Sequence subtractive hybridization and representational difference analysis were methods used for detection of gene expression studies and comparison of genome sequence repetitively [18]. Use of these methods was helpful for detection of human herpes simplex virus type 8 (HHV-8) [9, 19], GBV-A, GBV-B virus [20, 21], Tonovirus and norovirus [22].

Another independent approach is (SISPA) sequence-independent single-primer amplification circumvents used for detection of the unknown viral sequence by ligation of linker oligonucleotide sequence [23]. Further, it can be used for molecular cloning of viral genome for subsequent characterization. This method has been used successfully for the discovery of well-known Hepatitis E virus [10, 24] Parvovirus 2 and 3 [24] and Norwalk virus [11]. As viruses are devoid of consensus sequences, generally culture-based traditional and molecular biology-sequence-dependent and sequence independent technique are useful for the study of limited samples with limited output. Most of the viruses remain unidentified due to this reason.

Compared to above techniques metagenomics is the less biased approach. Any type of virus with either RNA or DNA as a genome, cultivable or uncultivable or novel viruses can be quickly detected. The word metagenomics denotes “transcendent” and “ome” is the all or every in Greek collectively means all genomic content. Metagenomics is the study of genetic material with the help of advanced genomic research technique’s and computational tools, directly from the environmental sample. Metagenomics approach bypasses the need for classical biochemical laboratory techniques for microbial analysis. With the help of metagenomics, one can investigate all types of genomic contents of a variety of organisms. This technique provided an indispensable tool for identification of nonculturable species of microbes. It is also used for investigation of known and culturable organisms with great accuracy. Another advantage to use this tool is it bypasses the need to isolate and culture individual species manually and the thereby it reduces the time required to study while providing more information. Initial metagenomics analysis of samples directly from raw environmental samples subsequently provides a necessary foundation for further lab-based analysis (**Table 1**). Metagenomics has been used for a variety of purposes, in diverse areas from the time of its discovery in 2002 when for the first time this approach was used in the virology field [12, 52].

### 4.3. Process of metagenomics

Metagenomics tool is a successful tool for surveillance in different environmental conditions such as freshwater, soil, marine water and gut of different organisms (**Table 1**) Recent advances in sequencing technology improved the speed of novel virus discovery and surveillance of environment [13, 53]. In 2000s, increase in literature related to metagenomics use in virome study and increase in a number of virus database show the ease of process. Recently government organization takes active participation in conducting surveillance programs [14, 15, 54, 55].

Basically, there are three main steps involved in metagenomics analysis of sample as follows:

1. Sample preparation
2. Sequencing
3. Bioinformatics analysis

| Year of study | Sample type   | Method of sequencing                         | Virus detected   | Reference |
|---------------|---|--|--|-----------|
| 2002          | Sea water   | Sanger's                                     |  | [12]      |
| 2003          | Feces   | Sanger's                                     |  | [16]      |
| 2004          | Marine sediments                                      | Sanger's                                     |  | [17]      |
| 2005          | Blood   | Sanger's                                     | Novel anellovirus  | [25]      |
| 2005          | Plasma  | SISPA  | Novel parvoviruses   | [18]      |
| 2005          | Nasopharyngeal aspirates                              | Sanger's                                     | Novel bocavirus  | [26]      |
| 2006          | Seawater  | Sanger's                                     | Novel RNA viruses  | [27]      |
| 2006          | Feces   | Sanger's                                     | Plant RNA viruses  | [28]      |
| 2007          | Honey bees  | 454 NGS                                      | Israeli acute paralysis virus  | [29]      |
| 2007          | Feces, urine, blood                                   | rolling circle amplification (RCA) technique | Novel polyomavirus   | [30]      |
| 2007          | Soil  | Sanger's                                     | Soil metagenomics overview   | [31]      |
| 2007          | Virioplankton   | Sanger's                                     | Virioplankton metagenome   | [32]      |
| 2008          | Feces   | Sanger's                                     | Study of diversity viruses in growing infants  | [33]      |
| 2008          | Feces   | Sanger's                                     | Novel picobirnavirus, picornavirus, norovirus and anellovirus, picornavirus, norovirus, picobirnavirus | [34]      |
| 2008          | Turkey feces  | 454NGS                                       | Novel bornavirus   | [35]      |
| 2008          | Hot spring water                                      | Sanger's                                     | Novel viruses in hot springs   | [36]      |
| 2008          | Bush kuru rat   | 454NGS                                       | Novel arenavirus   | [37]      |
| 2008          | Insect pool, skunk brain, human feces, sewer effluent | 454NGS                                       | Orthoreovirus and orbivirus  | [38]      |
| 2008          | SISPA   | Novel paralysis virus                        |  | [39]      |
| 2009          | plasma, liver biopsy                                  | 454NGS                                       | Novel LUJO virus   | [40]      |
| 2009          | grapevine   | 454NGS                                       | Novel marafivirus  | [41]      |
| 2009          | plant   | 454NGS                                       | Novel cucumovirus  | [41]      |
| 2009          | potable, reclaimed water                              | 454NGS                                       | Several animal and plant viruses   | [42]      |
| 2009          | Sea lion lungs  | Sanger's                                     | Novel California sea lion anellovirus  | [43]      |
| 2009          | Sea turtle swabs/tissues                              | 454NGS                                       | Novel sea turtle fibropapilloma virus  | [43]      |
| 2009          | Ant   | Sanger's                                     | <i>Solenopsis invicta</i> virus  | [44]      |
| 2009          | Feces   | 454NGS                                       | Klassevirus  | [45]      |
| 2009          | Plant   | Sanger's                                     | Sweet potatoes badnavirus and mastrevirus  | [46]      |
| 2010          | Brain   | 454NGS                                       | Astrovirus   | [45]      |
| 2010          | Feces   | 454NGS                                       | Novel chimpanzee associated circular virus   | [47]      |
| 2010          | Mosquitoes  | 454NGS                                       | Novel mycovirus  | [48]      |

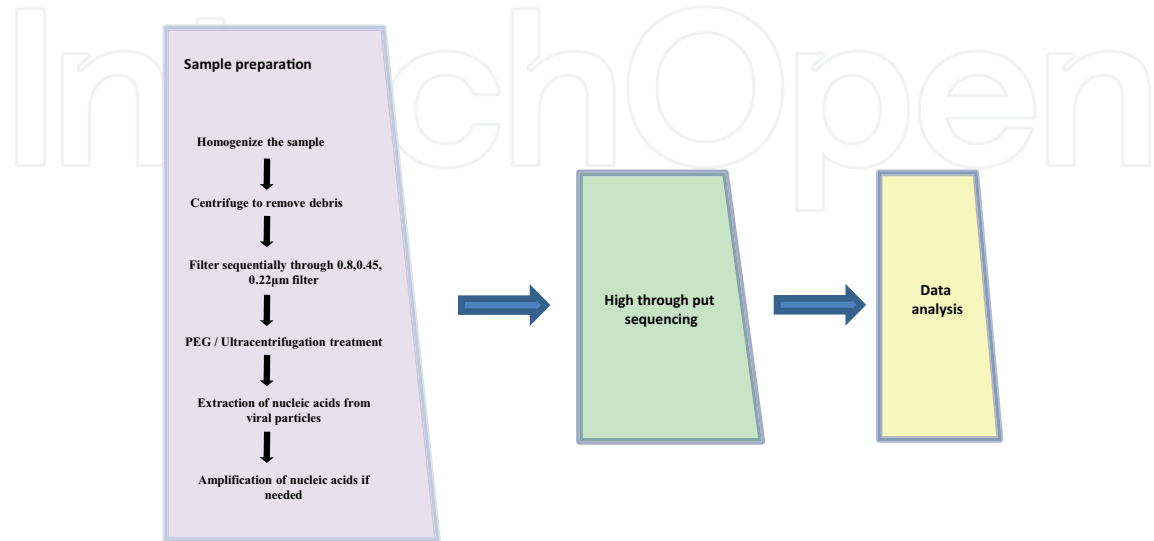


| Year of study | Sample type                       | Method of sequencing | Virus detected                                   | Reference |
|---------------|-----------------------------------|----------------------|--|-----------|
| 2011          | Plasma                            | 454NGS               | Novel simian hemorrhagic fever virus             | [49]      |
| 2011          | Feces                             | 454NGS               | Many novel species in pig: astrovirus, bocavirus | [50]      |
| 2011          | Liver, pancreas, intestine biopsy | 454NGS               | Novel turkey hepatitis virus                     | [51]      |

**Table 1.** Viruses discovered with metagenomics approach.

**1. Sample preparation and processing:** Since in metagenomics any type of sample can be analyzed with some pretreatment (or enrichment methods). However, for analysis of gut-associated virome collection of the different sample is done from different parts of the human gastrointestinal region. For accurate results, sample collection, proper handling, transportation, stage of the sample is very crucial. There are many standard protocols available for collection of different samples to laboratory and its storage techniques [37]. Different protocols are used for fluid sample and for tissue samples. The tissue sample is generally homogenized in autoclaved saline and collected supernatant filtered through 0.8, 0.45 and 0.2  $\mu\text{m}$  filters, this serial filtration procedure is used to separate larger particles and bacteria from viruses. See **Figure 1**.

There are different types of sample processing methods used earlier for extraction of viral genomic material [16, 56–58]. Based on studies done by many groups [56, 58–60], a framework designed by Shah et al. in 2014. A comparative analysis of three widely used sample processing methods for gut-associated RNA virome was done. The second processing method used in the separation of virus partials and DNA preparation gave good results. In that method, PEG treatment and ultracentrifugation steps are spatially separated by sonication step in PBS buffer to remove remnants of PEG. In this method based on bioinformatics



**Figure 1.** Overview of general procedure of metagenomics.

tools, like riboPicker tool version and blast of viral RNA sequence showed more number of virus domains present in the sample which were processed via the second method, while other methods showed more cellular noise [19].

2. **Sequencing:** The rate of metagenomics study was slow during Sanger sequencing when around 2005 other methods are yet to be evolved, Sangers sequencing was in use. Many studies in this period showed abundant diversity in viruses, analysis of human clinical samples also showed plenty of diversity, while speed of viral genome sequencing is increased several times during pyrosequencing. New viral communities of human and animals have been identified during this period. Some important discoveries are as follows: Astrovirus [21], Rhabdovirus [22], Coronavirus [23], Picornavirus [24], gammapaillomavirus [61]. This technology becomes popular in short time because of low cost, a high number of reads. This technology is also used for sequencing of the clinical sample from tissue fluids and tissue samples [11].

**Ion Torrent:** This is pH-based sequencing method with few steps are similar to pyrosequencing technology. Ion Torrent technology gives very rapid runs so it was very useful for targeted deletion of viral sequences from clinical samples such as HIV, HCV, polyomavirus, influenza virus, etc. This method was not a good choice for virologists for identification of new viruses because of low output.

**Illumina:** This technology is a high-throughput platform with low-cost rate of virus identification; many viruses from clinical samples have been identified using this technique.

**Pacific bioscience sequencing and nanopore sequencing:** These sequencing methods were not popular for metagenomics study because of high error rate [52].

3. **Bioinformatics analysis:** Bioinformatics analysis of raw sequence data generated from high-throughput sequencer is a critical step in novel virus discovery and even in diagnostics. There many ready to use pipelines available for analysis of raw data. VIP, VirFinder, Vipie, METAVIR, PHACCS, VIROME, HP Viewer, Fast virome Explorer, EzMAP, Vanator, viruspy and Viral\_genome\_annotator are few commonly used pipelines for viral metagenomics analysis. Typical workflow of viral metagenomics includes the following steps. Next-generation sequencing (NGS) data obtained is first subjected to trimming for removal of low-quality sequences and adaptor sequences, (Refer **Figure 2**). Second the trimmed data is subjected for removal of host (humans or bacteria) related sequences and third, these sequences are aligned to reference viral genomes for advance functional characteristics such as novel virus identification, viral taxonomy, identification of viral proteins and phylogenic analysis.

**Challenges involved in metagenomics:** For analysis of sequencing data of viral genome through high throughput, sequencing machine needs standard computational tools, software with a high accuracy of data analysis. This needs high-cost involvement with technical expertise. Few high-quality tools available for sequence data analysis such as Diamond [53], UBLAST [52] and Kaiju [54] have increased the speed of metagenomics study. Still, there is a need for technical improvement for rapid and accurate data analysis. The second challenge involved in data analysis of metagenomics sequencing is an assembly of the genome from thousands of small fragments. Assemblers used for the assembly of single genome sets during early times of sequencing study are outdated or non-useful for



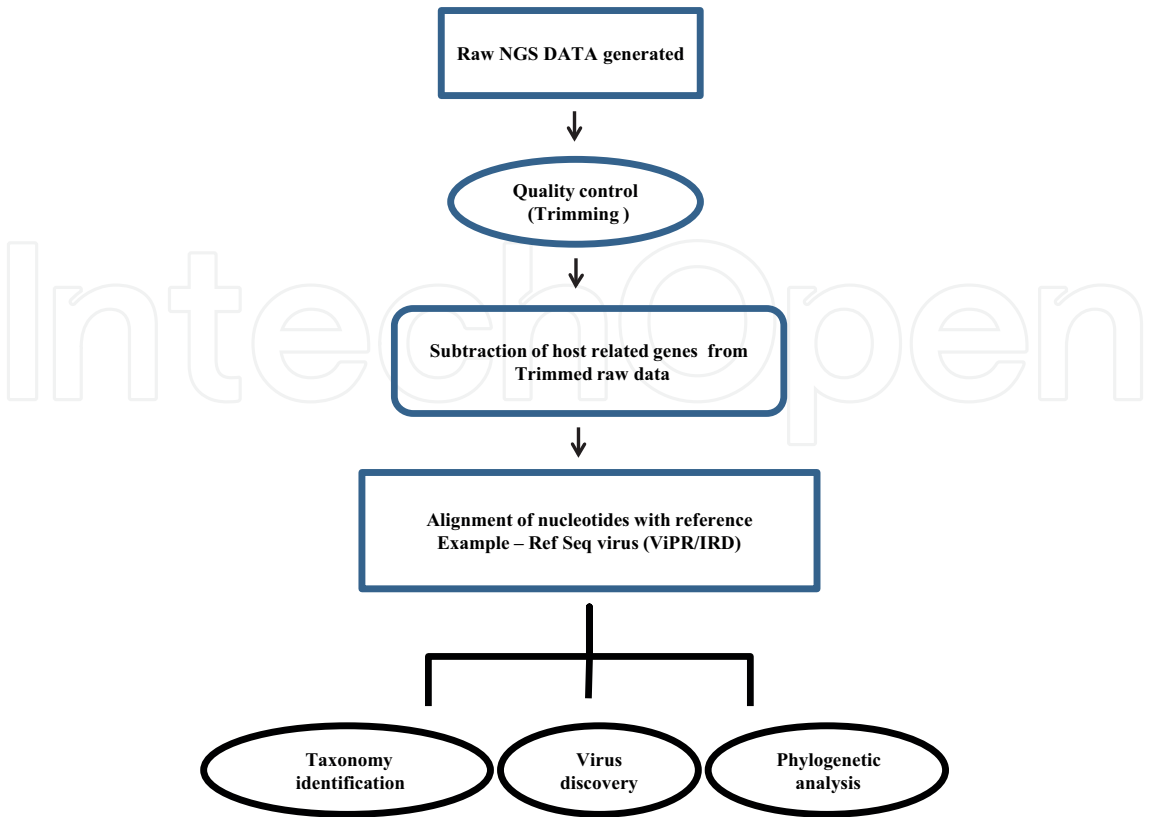


Figure 2. Workflow of metagenomics data analysis.

metagenomics; they create chimeric genomes which misinterpret the genome sequence. Now a days for such studies MetAMOS [55], Meta Velvet [62], MetaSPADes [57] assemblers are available. Still assembly process requires manual editing to sort out genomic chimeric generation [15]. Another challenge of virologists for data analysis is reference database deposited which sometimes may cause confusion or problems. If reference database is misinterpreted it will give a wrong interpretation of results. If reference database is high, it decreases the speed as a large number of sequence alignments are required to test data. Sequence data interpretation is a last and very decisive step for metagenomics. Still, we lack clear knowledge about the link between the diversity of virus in the environment and during outbreaks, our surveillance is merely based on a biased collection of only clinical samples and their study. This limits our knowledge about disease spread [63]. Prediction of future outbreaks and limiting the spread of disease needs proper study, development of strong tools [15] Therefore further extensive studies should be encouraged for obtaining maximum and precise knowledge of environmental and gut-associated virome.

4.4. Applications in gut-associated virome analysis

1. **Epidemic and endemic surveillance:** Several reports of unknown pathogenic virus outbreaks in history suggest the need for comprehensive study of virus-host interaction during disease and disease-causing viruses is a big threat to the human population. Well, known examples of zoonotic virus transmission are Nipah virus from fruit bats [58] and Ebola virus from bushmeat [60]. This creates a need for continuous surveillance of diseases in the community. David et al. in 2017 [15] gave a comprehensive explanation about disease outbreak and its diagnosis with

the help of surveillance pyramid. The surveillance pyramid explains during disease spread in the community only a few diagnosed cases are reported, the individuals carrying symptoms of the disease and the carriers of the disease are not reported. This phenomenon creates biasedness in sampling. Therefore metagenomics study has been proved a useful tool for constant surveillance of gastrointestinal tract pathogenic virome community. As well as some endemic viral diseases, which causes common gastrointestinal health concerns in community, e.g., astrovirus, calicivirus, norovirus, and torovirus [64], herpesviruses, hepatitis E virus, epstein bar virus, coxsackieviruses, and surveillance with the metagenomics study is useful.

2. **Discovery of new viruses and classification:** Metagenomics is a powerful tool for identification of novel organism(s). Screening of different gut samples can be useful to study novel gut-associated viruses. Initially with the sequence-based studies of Markel cell carcinoma new *human papillomavirus* has been identified. Markel cell carcinoma is human skin tissue carcinoma, where virus DNA found to be integrated into tumor tissue [65]. Subsequent studies have revealed the diversity of gut-associated viruses in different animals which help in the study of past zoonotic occurred in history. Human-rodent's interaction is well known due to civilization in forest areas or due to the domestication of animals this is leading cause of zoonotic outbreaks. Knowledge of outbreaks in past and monitoring of the present status of the spread of known pathogenic viruses and closely associated pathogenic human viruses provides a base to predict future outbreaks. This approach is also useful to limit the epidemiology of recurrent outbreaks with the study of disease-prone viruses and characterization of unknown viruses. Phan et al. in 2011 extensively studied fecal sample from wild rodents in Virginia and they characterized viruses belonging to mammalian virus families, many new viral families, two new genera were identified. Two viruses closely related to *Aichivirus*, an associated with acute gastroenteritis worldwide, were characterized through the study [66].

Turkey meat is very popular in the USA and its production is an important part of US economy. One study conducted in California in March 2011 on turkey which was suffering from turkey viral hepatitis. Pyrosequencing of RNA, extracted from liver revealed the presence of novel picornaviruses named as *turkey hepatitis virus* [51]. Another study on cattle's suffering from the unknown disease in Germany and Netherlands affected milk production. Metagenomics study discovered the new virus, *Schmallenberg virus*, from infected cow sample [67]. Identification and characterization of such viruses will help in facing problems which have a negative impact on countries economic status. Similar to domestic animals, wild-type animals can also act as a reservoir of novel pathogens. Two novel simian hemorrhagic fever viruses diverse from original simian hemorrhagic fever virus were identified from African green monkeys. *Simian hemorrhagic fever virus* has not yet found to infect human but clinical indices comparable with human *Ebola* and *Marburg viruses*. This analogy makes it in the suspect list of emerging viruses [49].

3. **Diagnostic Metagenomics** is a potent method that allows broad analysis of relative genetic variation among viruses and can be used for the study of host-pathogen interactions. This is also more popular because it can be used for uncultivable organisms as well. The recently rising approach is to use metagenomics during epidemics and outbreaks, with a given large number of samples in a lesser time. In *hepatitis C virus* (HCV) infection, identification of infection is a challenging task due to lack of apparent symptoms and lack of easy laboratory tests for differentiation of acute and chronic phase of the disease. Available

molecular methods for virus diagnostic purpose are tedious, time-consuming and costly. A recent report from Escobar-Gutierrez et al. described the use of next-generation sequencing (NGS) method in the diagnosis of HCV infection. NGS allows cost-effective analysis of a large number of samples in detail. The study showed low-frequency mutations, genetic variation [68]. Genetic shift and re-assortment viruses are a leading cause of the emergence of a new strain of viruses, especially in RNA viruses. Well a known example is influenza virus, many pandemics and deaths in history. The recent H1N1 virus is a combination of swine, human and avian genomic segments of RNA [69]. The best approach of metagenomics study in 2009 H1N1 pandemic is the use of metagenomics for characterization and detail study of the virus, followed by manufacture of microarray-based virochip for rapid detection and differential screening from seasonal virus [70].

4. **Evolution of host-virus interaction:** Evolution of RNA viruses is comparatively fast process than DNA viruses. Study of evolution is necessary to understand the source of new variance, spread and keep a check on epidemic initiating variant. In emerging RNA virus, *norovirus* causative agent of gastroenteritis inter-host, intra-host, and transmission of the new variant has been studied. Usually, it is a self-limiting acute disease but in immune-compromised individuals and in newborns it may cause morbidity and mortality. No vaccine or drugs are available for treatment. A report from Bull et al. hypothesized based on metagenomics study that, *norovirus* has multiple mechanisms of evolution. Chronic hosts are a major reservoir of new variants while acute patients generally possess a single variant. NGS approach for use assists in comprehensive study of viral population dynamics [71]. Characterization of cardiovirus genus originally believed to possess two genera, metagenomics study has revealed five new genera with full characterization. Cardioviruses are the causative agent of enteric diseases in mice with multiple symptoms. In humans, it causes encephalitis-like condition and diarrhea in children's [72]. Metagenomics based studies help in designing future approach with these new genotypes and associated diseases.

## 5. Conclusion

The metagenomics studies have a huge potential to describe about diversity of microbiome in gut microflora and most importantly directly in infectious samples. Among all pathogens viruses are the ones, who cause severe illness to mankind. With rapid improvement in the genomic sequencing techniques, the overall metagenomics approach is very valuable for discovery of new viruses, novel genes, surveillance of pathogens, discover new pathway, host virus interaction, functional studies. The leads obtained through this exercise may have great impact on early diagnosis and treatment. While metagenomic studies also experience limitations and challenges, which need to overcome in near future to obtain a precise results. Unified genomic extraction techniques and development of improved analysis modules may suffice the needs of metagenomics in future.

## Conflict of interest

Authors declare no conflict of interest.

## Author details

Prudhvi Lal Bhukya<sup>1\*</sup> and Renuka Nawadkar<sup>2</sup>

\*Address all correspondence to: saiprudhvi21@gmail.com

1 Department of Biotechnology, Krishna University, Machilipatnam, Andhra Pradesh, India

2 Yashwantrao Chavan College of Science, Karad, India

## References

- [1] Hobbie JE, Daley RJ, Jasper S. Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Applied and Environmental Microbiology*. 1977;**33**:1225-1228
- [2] Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: Discovery and emergence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2012;**367**:2864-2871
- [3] Salim AF, Phillips AD, Farthing MJ. Pathogenesis of gut virus infection. *Baillière's Clinical Gastroenterology*. 1990;**4**:593-607
- [4] Brogden KA, Guthmiller JM. *Polymicrobial Diseases*. Washington, D.C.: ASM Press; 2002
- [5] Rene E, Verdon R. Upper gastrointestinal tract infections in AIDS. AIDS GIT group. *Baillière's Clinical Gastroenterology*. 1990;**4**:339-359
- [6] Solonenko SA, Sullivan MB. Preparation of metagenomic libraries from naturally occurring marine viruses. *Methods in Enzymology*. 2013;**531**:143-165
- [7] Draghici S, Khatri P, Eklund AC, Szallasi Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends in Genetics*. 2006;**22**:101-109
- [8] Balsalobre-Arenas L, Alarcon-Cavero T. Rapid diagnosis of gastrointestinal tract infections due to parasites, viruses, and bacteria. *Enfermedades Infecciosas y Microbiología Clínica*. 2017;**35**:367-376
- [9] Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*. 1994;**266**:1865-1869
- [10] Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010;**466**:334-338
- [11] Quan PL, Firth C, Conte JM, Williams SH, Zambrana-Torrel CM, Anthony SJ, et al. Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;**110**:8194-8199
- [12] Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;**99**:14250-14255

- [13] Ansorge WJ. Next-generation DNA sequencing techniques. *New Biotechnology*. 2009;**25**:195-203
- [14] Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;**486**:207-214
- [15] Nieuwenhuijse DF, Koopmans MP. Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Frontiers in Microbiology*. 2017;**8**:230
- [16] Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*. 2003;**185**:6220-6223
- [17] Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F. Phage community dynamics in hot springs. *Applied and Environmental Microbiology*. 2004;**70**:1633-1640
- [18] Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. New DNA viruses identified in patients with acute viral infection syndrome. *Journal of Virology*. 2005;**79**:8230-8236
- [19] Shah JD, Baller J, Zhang Y, Silverstein K, Xing Z, Cardona CJ. Comparison of tissue sample processing methods for harvesting the viral metagenome and a snapshot of the RNA viral community in a Turkey gut. *Journal of Virological Methods*. 2014;**209**:15-24
- [20] Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, et al. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;**92**:3401-3405
- [21] Quan PL, Wagner TA, Briese T, Torgerson TR, Hornig M, Tashmukhamedova A, et al. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerging Infectious Diseases*. 2010;**16**:918-925
- [22] Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe JJ, et al. A novel rhabdovirus associated with acute hemorrhagic fever in Central Africa. *PLoS Pathogens*. 2012;**8**:e1002924
- [23] Honkavuori KS, Briese T, Krauss S, Sanchez MD, Jain K, Hutchison SK, et al. Novel coronavirus and astrovirus in Delaware Bay shorebirds. *PLoS One*. 2014;**9**:e93395
- [24] Boros A, Nemes C, Pankovics P, Kapusinszky B, Delwart E, Reuter G. Identification and complete genome characterization of a novel picornavirus in Turkey (Meleagris gallopavo). *The Journal of General Virology*. 2012;**93**:2171-2182
- [25] Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *BioTechniques*. 2005;**39**:729-736
- [26] Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;**102**:12891-12896
- [27] Culley AI, Lang AS, Suttle CA. Metagenomic analysis of coastal RNA virus communities. *Science*. 2006;**312**:1795-1798
- [28] Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, et al. RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biology*. 2006;**e3**:4



- [29] Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*. 2007;**318**:283-287
- [30] Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, et al. Identification of a third human polyomavirus. *Journal of Virology*. 2007;**81**:4130-4136
- [31] Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology*. 2007;**73**:7059-7066
- [32] Bench SR, Hanson TE, Williamson KE, Ghosh D, Radosovich M, Wang K, et al. Metagenomic characterization of Chesapeake Bay viroplankton. *Applied and Environmental Microbiology*. 2007;**73**:7629-7641
- [33] Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, et al. Viral diversity and dynamics in an infant gut. *Research in Microbiology*. 2008;**159**:367-373
- [34] Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: Viral detection and discovery. *PLoS Pathogens*. 2008;**4**:e1000011
- [35] Honkavuori KS, Shivaprasad HL, Williams BL, Quan PL, Hornig M, Street C, et al. Novel Bornavirus in psittacine birds with proventricular dilatation disease. *Emerging Infectious Diseases*. 2008;**14**:1883-1886
- [36] Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D. Assembly of viral metagenomes from yellowstone hot springs. *Applied and Environmental Microbiology*. 2008;**74**:4164-4174
- [37] Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *The New England Journal of Medicine*. 2008;**358**:991-998
- [38] Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL. Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathogens*. 2008;**4**:e1000163
- [39] Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, et al. Viral genome sequencing by random priming methods. *BMC Genomics*. 2008;**9**:5
- [40] Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathogens*. 2009;**5**:e1000455
- [41] Al Rwahnih M, Daubert S, Golino D, Rowhani A. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology*. 2009;**387**:395-401
- [42] Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*. 2009;**11**:2806-2820
- [43] Ng TF, Suedmeyer WK, Wheeler E, Gulland F, Breitbart M. Novel anellovirus discovered from a mortality event of captive California Sea lions. *The Journal of General Virology*. 2009;**90**:1256-1261
- [44] Valles SM, Hashimoto Y. Isolation and characterization of *Solenopsis invicta* virus 3, a new positive-strand RNA virus infecting the red imported fire ant, *Solenopsis invicta*. *Virology*. 2009;**388**:354-361



- [45] Greninger AL, Runckel C, Chiu CY, Haggerty T, Parsonnet J, Ganem D, et al. The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology Journal*. 2009;**6**:82
- [46] Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, et al. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*. 2009;**388**:1-7
- [47] Blinkova O, Victoria J, Li Y, Keele BF, Sanz C, Ndjango JB, et al. Novel circular DNA viruses in stool samples of wild-living chimpanzees. *The Journal of General Virology*. 2010;**91**:74-86
- [48] Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, Lentz SM, et al. Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. *PLoS Neglected Tropical Diseases*. 2010;**4**:e878
- [49] Lauck M, Hyeroba D, Tumukunde A, Weny G, Lank SM, Chapman CA, et al. Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One*. 2011;**6**:e19056
- [50] Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E. The fecal virome of pigs on a high-density farm. *Journal of Virology*. 2011;**85**:11697-11708
- [51] Honkavuori KS, Shivaprasad HL, Briese T, Street C, Hirschberg DL, Hutchison SK, et al. Novel picornavirus in Turkey poults with hepatitis, California, USA. *Emerging Infectious Diseases*. 2011;**17**:480-487
- [52] Kumar A, Murthy S, Kapoor A. Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Research*. 2017;**239**:172-179
- [53] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 2015;**12**:59-60
- [54] Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*. 2016;**7**:11257
- [55] Treangen TJ, Koren S, Sommer DD, Liu B, Astrovskaya I, Ondov B, et al. MetAMOS: A modular and open source metagenomic assembly and analysis pipeline. *Genome Biology*. 2013;**14**:R2
- [56] Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nature Protocols*. 2009;**4**:470-483
- [57] Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: A new versatile metagenomic assembler. *Genome Research*. 2017;**27**:824-834
- [58] Yob JM, Field H, Rashdi AM, Morrissy C, van der Heide B, Rota P, et al. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. *Emerging Infectious Diseases*. 2001;**7**:439-441
- [59] Hurwitz BL, Deng L, Poulos BT, Sullivan MB. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology*. 2013;**15**:1428-1440

- [60] Mann E, Streng S, Bergeron J, Kircher A. A review of the role of food and the food system in the transmission and spread of Ebolavirus. *PLoS Neglected Tropical Diseases*. 2015;**9**:e0004160
- [61] Phan TG, Vo NP, Aronen M, Jartti L, Jartti T, Delwart E. Novel human gammapapilloma-virus species in a nasal swab. *Genome Announcements*. 2013;**1**:e0002213
- [62] Afiahayati SK, Sakakibara Y. MetaVelvet-SL: An extension of the velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Research*. 2015;**22**:69-77
- [63] La Rosa G, Libera SD, Iaconelli M, Ciccaglione AR, Bruni R, Taffon S, et al. Surveillance of hepatitis a virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. *BMC Infectious Diseases*. 2014;**14**:419
- [64] Tran A, Talmud D, Lejeune B, Jovenin N, Renois F, Payan C, et al. Prevalence of rotavi-rus, adenovirus, norovirus, and astrovirus infections and coinfections among hospital-ized children in northern France. *Journal of Clinical Microbiology*. 2010;**48**:1943-1946
- [65] Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. 2008;**319**:1096-1100
- [66] Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. *PLoS Pathogens*. 2011;**7**:e1002218
- [67] Hoffmann B, Scheuch M, Hoper D, Jungblut R, Holsteg M, Schirrmeier H, et al. Novel orthobunyavirus in cattle, Europe, 2011. *Emerging Infectious Diseases*. 2012;**18**:469-472
- [68] Escobar-Gutierrez A, Vazquez-Pichardo M, Cruz-Rivera M, Rivera-Osorio P, Carpio-Pedroza JC, Ruiz-Pacheco JA, et al. Identification of hepatitis C virus transmission using a next-generation sequencing approach. *Journal of Clinical Microbiology*. 2012;**50**:1461-1463
- [69] Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, et al. Antigenic and genetic characteristics of swine-origin 2009 a(H1N1) influenza viruses circulating in humans. *Science*. 2009;**325**:197-201
- [70] Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, et al. A metage-nomic analysis of pandemic influenza a (2009 H1N1) infection in patients from North America. *PLoS One*. 2010;**5**:e13381
- [71] Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of intra-and interhost dynamics to norovirus evolution. *Journal of Virology*. 2012;**86**:3219-3229
- [72] Blinkova O, Kapoor A, Victoria J, Jones M, Wolfe N, Naeem A, et al. Cardioviruses are genetically diverse and cause common enteric infections in south Asian children. *Journal of Virology*. 2009;**83**:4631-4641

